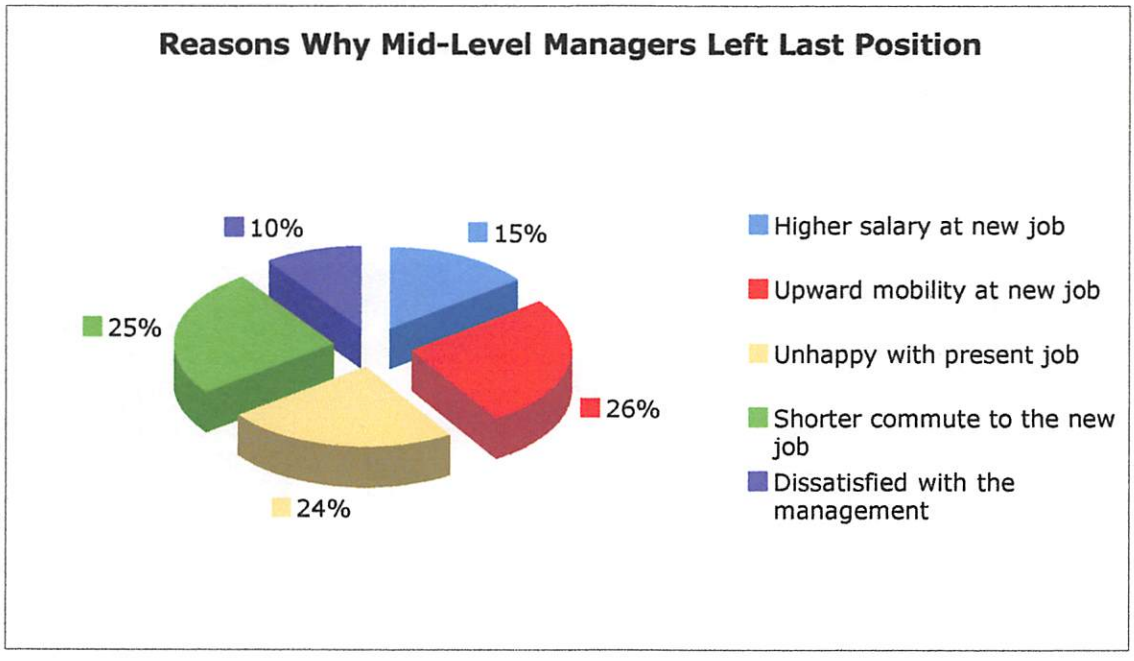
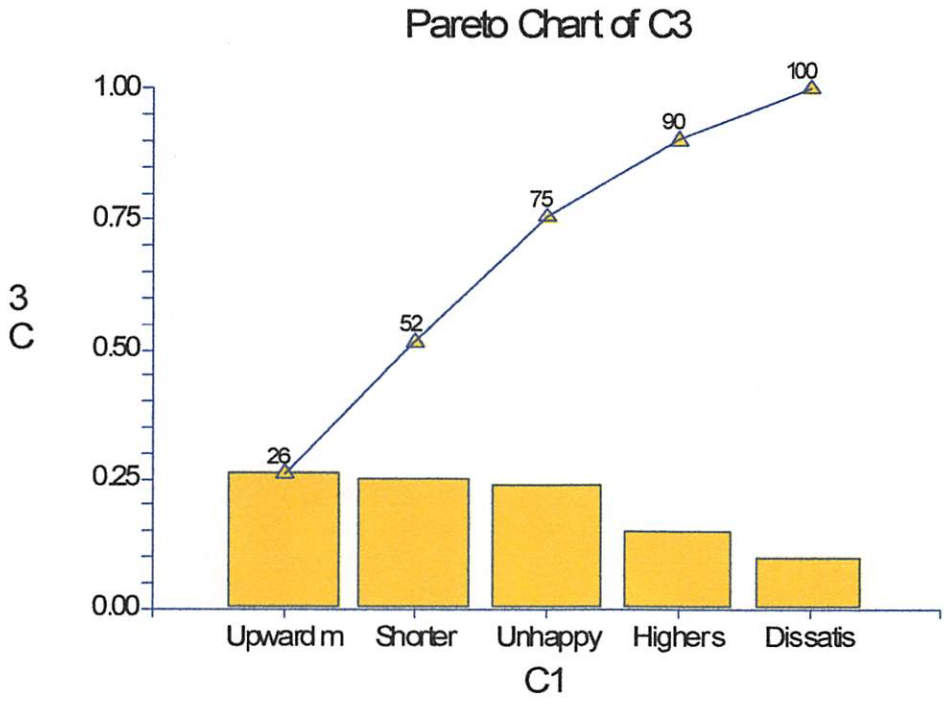
Joe Bustillos

Data Analysis & Interpretation Homework Assignment
ED734a - Cadre X - Summer 2005

## Problem 1: Mid-Level Managers




The three top reasons given for resigning from their most recent position are: upward mobility, shorter commute and unhappiness with present job. The top three reasons also represent 75% of the reasons for leaving. Only three

percentage points separate the top three choices. One thing that I find most surprising is that the second lowest reason is "higher salary." "Upward mobility most likely includes higher salary but means much more than just dollars.

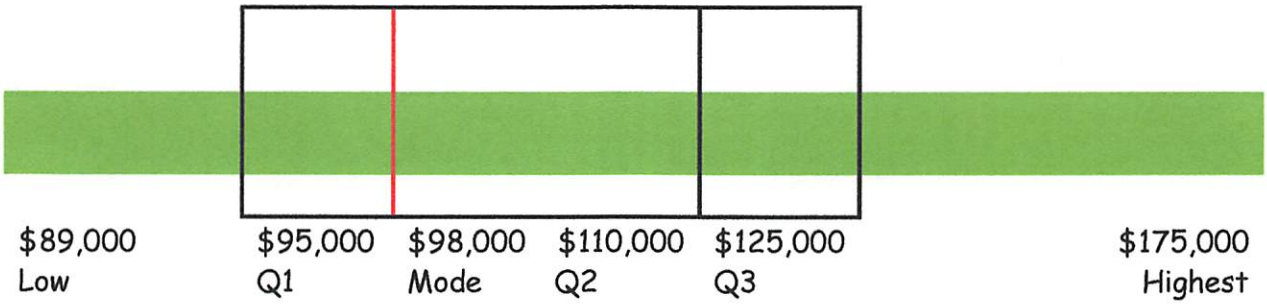
**Problem 2: Household Income & Home Prices**

The Mode prices of the homes leads one to believe that most of the homes are around $278K, making it the most common price, but there are enough homes significantly more expensive to skew the Mean and Median prices to the $458K and $492K prices respectively. Income-wise the Mean, Mode and Median are in greater agreement with a much less range of differences. But the difference between income and range of home prices may lead one to believe that a number of families in the neighborhood are living at the edge if not beyond their means. Further research might indicate that the values of certain homes (or sections of the neighborhood) might have shot up in value after the family moved in. Conversely, the property value of certain homes (or sections of the neighborhood) may have significantly dropped. The rise or fall in prices might explain the difference in range of home prices while the income levels remained more aligned.

Excellent answer, but Dr. Madjidi likes you to also put the numbers into a practical explanation. For example, The average price of a home was $458,000 with a standard deviation of $58,000, and the most frequently occurring price was $278,000. Exactly half of the home prices fell above $492,000 and exactly helf fell below this same number.

This putting the numbers into simple terms also applies to the salary information.

**Problem 3: IT Salaries**




a. Interquartile range = $30,000
b. Range = $86,000
c. Possible skew = the data is skewed to the right or positively skewed To the left, or positively skewed
d. 73rd percentile is $122,000, which is more than 73 out of 100 IT managers.
e. 50th percentile is $110,000, with half the salaries above and the other half below.

## Problem 4: Price of homes

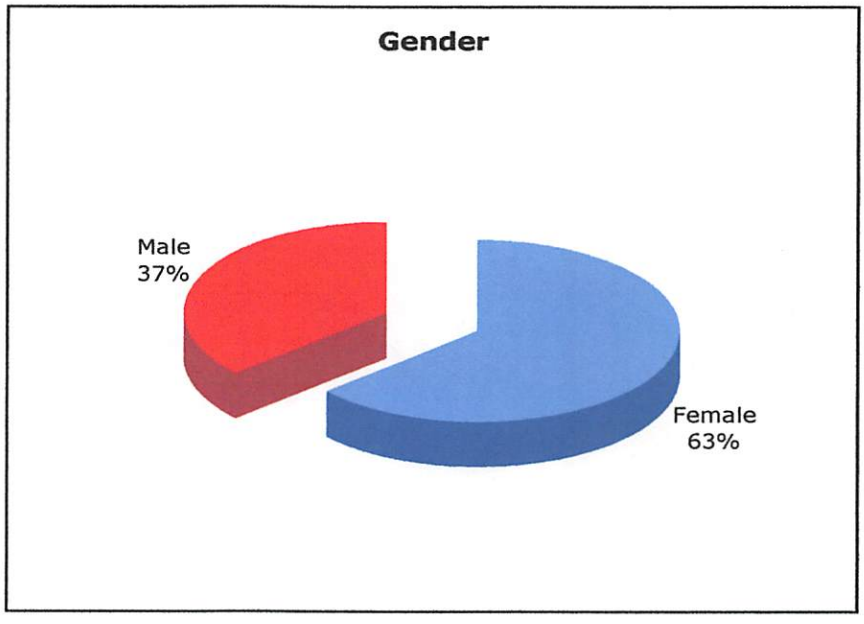
a. What percentage of homes cost between $250,000 and $650,000? **95.4%**
b. What percentage of the homes cost more than $750,000? **.13%**
c. What percentage of the homes cost between $350,000 and $750,000? **83.99%**
d. What percentage of the homes cost less than $650,000? **97.73%**
e. Approximate the percentage of the homes that cost between $458,000 and $550,000? **34%** .4772 + .500 = 30.94%

## Problem 5: Customer Support Center

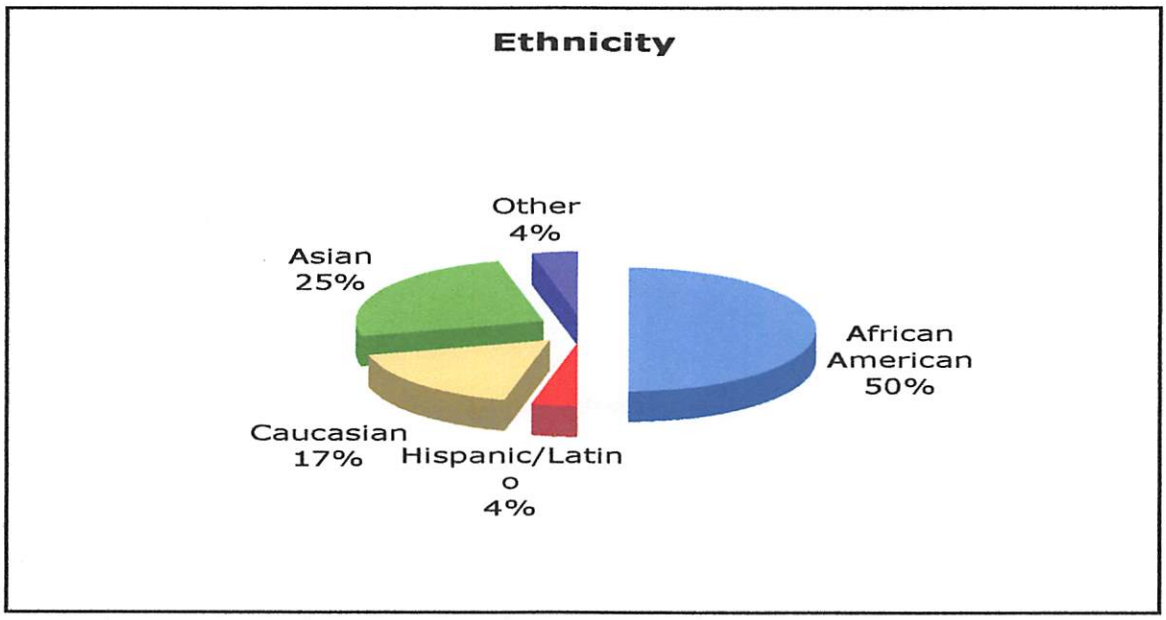
a. What is the probability that the center receives less than 195 calls a day? **2.27%**
b. What is the probability that the center receives more 215 calls a day? **15.87%**
   34.13 + .5 = 84.13%
c. What is the probability that the center receives between 195 and 255 calls a day? **81.85%**
d. What is the probability that the center receives between 215 and 255 calls a day? **68.26%**
e. What is the probability that the center receives over 295 calls a day? **.13%**

## Problem 6: LMUENGR Survey

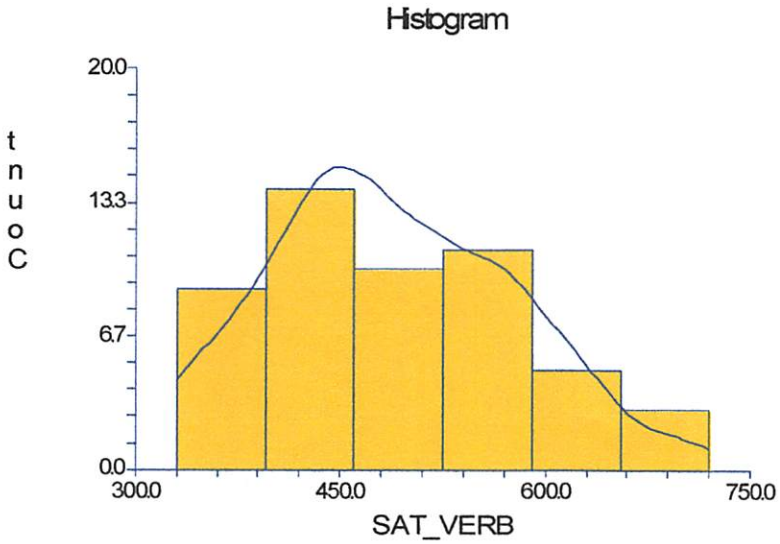
There were 52 students in the student with 33 students (approximately 63%) females and 19 students (approximately 37%) males. In the Ethnicity category 50% of the students were African American. The next largest group (25%) were Asian students.



| | SAT Math | SAT Verbal |
|---|---|---|
| Mean | 577.11 | 488.27 |

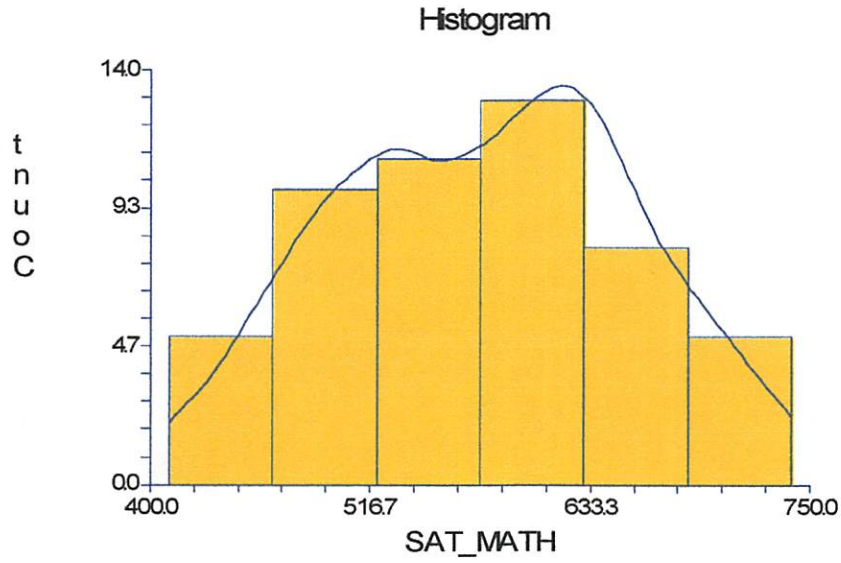



| | | |
|---|---|---|
| Mode | 610 | |
| Median | 575 | 480 |
| SD | 79.24 | 90.92 |
| Min | 410 | 330 |
| Max | 740 | 720 |
| Range | 330 | 390 |
| 1st Quartile | 520 | 422.5 |
| 3rd Quartile | 637.5 | 555 |




The students scored lower on the SAT Verbal than on the SAT Math (engineering students, right?). The median and mode SAT Math scores were 95 and 89 points higher, respectively, than SAT Verbal scores. Although the mean, mode, median scores were higher on the SAT Math, the SAT Verbal's range was greater.

SAT Verbal had an average of 488.2, with an SD of 90.9. Half of the students were higher and half lower than 480 (the median). The scores were between 330 and 720 with a range of 390.
SAT Math had an average of 577, with an SD of 79.24. Half of the students were higher and half lower than 575 (the median). The most common score was 610, and the scores fell between 410 and 740 with a range of 330.




**Problem 7: Concern about School Violence**

a. State the problem: reported increase in crime in a California school district.
Safety at schools
b. State a purpose for a study consistent with the scenario above: This study will be conducted in order to determine if there is a difference in the concern regarding the reported number of crime incidents based on family household incomes, ethnicity, the level of education of the head of the household, the number of school aged children in the family, or combination of any of the factors.
c. List at least five research questions consistent with purpose in item b:
   - What is the actual number of incidents over the course a school year, going back at least ten years (if possible) in order to track the trend of incidents in this particular school?
   - What is the difference between the reported trend of the district and this particular school?
   - What is the perception of crime incidents (No Concern, Low Concern, Neutral, High Concern, Extreme Concern) in local households based on family household incomes? What is the actual number of involvement in crime incidents based on family household incomes?
   - What is the perception of crime incidents (No Concern, Low Concern, Neutral, High

Concern, Extreme Concern) in local households based on ethnicity? What is the actual number of involvement in crime incidents based on ethnicity?

- What is the perception of crime incidents (No Concern, Low Concern, Neutral, High Concern, Extreme Concern) in local households based on the education of the head of the household? What is the actual number of involvement in crime incidents based on the education of the head of the household?
- What is the perception of crime incidents (No Concern, Low Concern, Neutral, High Concern, Extreme Concern) in local households based on the number of school aged children in the family? What is the actual number of involvement in crime incidents based on the number of school aged children in the family?
- What is the perception of crime incidents (No Concern, Low Concern, Neutral, High Concern, Extreme Concern) in local households based on family household incomes? What is the actual number of involvement in crime incidents based on family household incomes?

d. Identify the variables in the study and state at what level (numeric or attribute) would you measure each variable. Give examples.
   - Family household incomes – numeric attribute ($20,000 and below, $21,000 - $50,000, $51,000 - $80,000, $81,000 – $100,000, $101,000 - $130,000, $131,000 and above.)
   - Ethnicity – attribute (Caucasian, African American, Latino/a, Native American, Asian, Other)
   - Level of education in household – attribute (High School Diploma, Associate Degree, Bachelor's Degree, Master's Degree, Professional Degree, Doctorate)
   - Number of crime incidents – numeric
   - Number of school aged children in the family – numeric attribute (1, 2, 3, 4, 5 or more)
   - Level/Degree of concern of crime – attribute (No Concern, Low Concern, Neutral, High Concern, Extreme Concern)

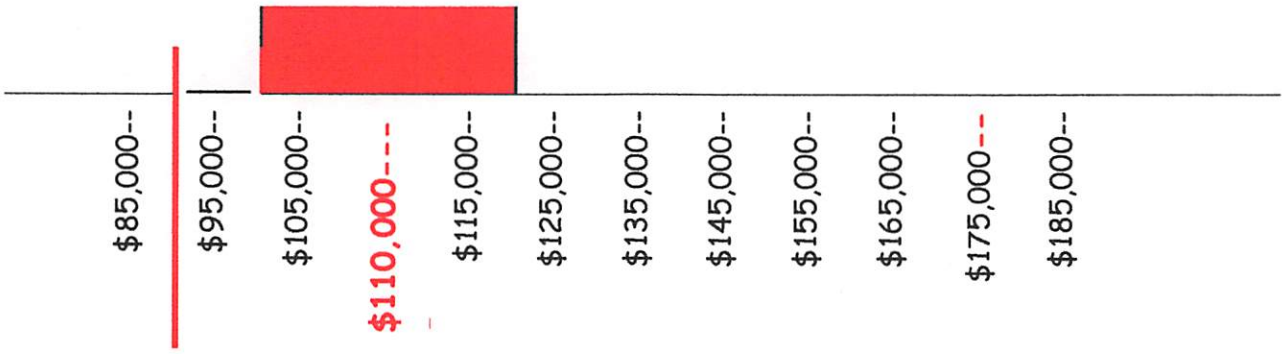
**Problem 8: Internet Training**

a. What is the probability that a student has completed the course in less than 11 hours? 90/140 or 64%
b. What is the probability that a student has completed the course who is older than 40? 60/140 or 43%
c. What is the probability that a student has completed the course in less than 5 hours and is over 18? 15/140 or 11%
d. What is the probability that a student has completed the course in over 5 hours and is below 18? 40/140 or 29%
e. What is the probability that a student has completed the course in 10 hours or less, given the student is 18-40? 15/35 or 43%
f. What is the probability that a student is 18-40 years old given the student has

completed the course in 11 or more hours? 20/50 or 40%

g. What is the probability that a student has completed the course in less than 5 hours or is below 18 years of age? 60/140 or 43%
h. What is the probability that a student has completed the course in less than 11 hours or is older 18? 120/140 or 86%

**Problem 9: IT Managers @ 100 high tech companies**




IQR = $30,000

There is a skew to the right or a positive skew with the 81% of the salaries at or below the median salary of $110,000.

**Problem 10: Medical Center Parking**

a. What is the problem statement: there is a shortage of parking spaces.
b. What are the objectives of the study:
    - To determine if a parking structure would be profitable.
    - To determine if patients will pay to park.
c. State several research questions:
    - Is there a difference between genders and who is willing to pay to park?
    - Is there a difference between age and who is willing to pay to park?
    - Is there a difference between paying to park and the annual income of patients?
d. Identify the variables and their level of measurements (Attr, or Num)
    - Gender – attribute
    - Age – attribute
    - Income – numeric
    - Amount willing to pay – numeric
    - Paying $3 to park - numeric
e. Is there anything you could improve on in this design? How?

- Why are you only surveying people on Tuesdays? Other days not make enough money to pay for the structure.
- Why are you asking gender and age? Will that help with demographics in other studies?
- Categorize annual incomes rather than leaving it open-ended.